Available online at www.sciencedirect.com
SCIENCE DIRECT®
Veterinary Immunology and Immunopathology 105 (2005) 235–245
Veterinary immunology and immunopathology
www.elsevier.com/locate/vetimm

A 7872 cDNA microarray and its use in bovine functional genomics

Robin E. Everts [a], Mark R. Band [b], Z. Lewis Liu [a,1], Charu G. Kumar [a,b], Lei Liu [b], Juan J. Loor [a], Rosane Oliveira [a], Harris A. Lewin [a,*]

[a] *Department of Animal Sciences, University of Illinois, 206 Edward R. Madigan Laboratory, 1201 West Gregory Drive, Urbana, IL 61801, USA*
[b] *W.M. Keck Center for Comparative and Functional Genomics, University of Illinois, 1201 West Gregory Drive, Urbana, IL 61801, USA*

Abstract

The strategy used to create and annotate a 7872 cDNA microarray from cattle placenta and spleen cDNA sequences is described. This microarray contains approximately 6300 unique genes, as determined by BLASTN and TBLASTX similarity search against the human and mouse UniGene and draft human genome sequence databases (build 34). Sequences on the array were annotated with gene ontology (GO) terms, thereby facilitating data analysis and interpretation. A total of 3244 genes were annotated with GO terms. The array is rich in sequences encoding transcription factors, signal transducers and cell cycle regulators. Current research being conducted with this array is described, and an overview of planned improvements in our microarray platform for cattle functional genomics is presented.

Keywords: Microarray; cDNA; Annotation; Cattle

1. Introduction

Microarray technology has transformed experimental biology by enabling quantitative analysis of the transcriptome of cells and tissues (Schena et al., 1995). Transcript profiles obtained using microarrays have been useful for elucidating metabolic pathways in healthy and diseased tissues (Hansson et al., 2004), identifying critical pathways in animal development (Hamatani et al., 2004) and predicting complex traits such as animal behavior (Whitfield et al., 2003). Practical applications derived from the global accessibility of the transcriptome are the classification of transformed cell-types (Golub et al., 1999; Takahashi et al., 2001) and the identification of novel targets for drugs and vaccines (Marton et al., 1998; Lotinun et al., 2003). In veterinary medicine and the animal sciences, gene expression profiling has the potential to

* Corresponding author. Tel.: +1 217 333 5998; fax: +1 217 244 5617.
E-mail address: h-lewin@uiuc.edu (H.A. Lewin).
[1] Current address: National Center for Agricultural Utilization Research, USDA-ARS, 1815 North University Avenue, Peoria, IL 61614, USA.

0165-2427/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.vetimm.2005.02.003

dramatically improve our understanding of physiology in health and disease, especially for diseases that lack an appropriate animal model. Significant applications in disease diagnosis are also on the horizon, as well as improved feed formulations for animals of different ages and physiological states, increased efficiency of assisted reproduction and enhanced rate of genetic improvement.

The development of microarray technology for cattle functional genomics closely paralleled advances made in low cost, high throughput DNA sequencing and the commercial availability of robotic instruments for spotting small volumes of DNA. The first reported cattle cDNA microarray was relatively small, containing ~700 unique genes (Yao et al., 2001; Burton et al., 2001). A 3800-element cattle cDNA microarray created at the University of Illinois (Band et al., 2002) was the first large-scale microarray available for functional genomic studies. Other cattle cDNA microarrays have been produced (Hashizume et al., 2002; Ishiwata et al., 2003; Herath et al., 2004), including the 18,000 cDNA element array produced by the Bovine Functional Genomics Consortium (Suchyta et al., 2003b). Studies performed on cattle have been diverse, ranging from gene-expression profiling of different tissues (Band et al., 2003; Suchyta et al., 2003a) to the analysis of beef quality (Reverter et al., 2003), gestation and fetal development (Ishiwata et al., 2003; Herath et al., 2004) and immune responses (Burton et al., 2001; Coussens et al., 2002). From the limited number of publications to date describing the use of cattle microarrays (13 cited in *PubMed* since 2001), it is apparent that the application of gene expression profiling to problems in bovine biology and medicine has only recently begun. In this article, we describe the current state-of-the-art in microarray technology and gene annotation for cattle functional genomics at the University of Illinois, briefly summarize ongoing collaborative research in animal health, nutrition, development and genetics, and present our plans for technology improvements.

2. Microarray construction

A collection of 12,620 ESTs from a normalized and subtracted cattle placenta cDNA library and 6144 ESTs from a normalized and subtracted spleen library (http://titan.biotec.uiuc.edu/cattle/cattle_project.htm) were used to select new cDNA inserts to be added to a previously described 3800-element array (Band et al., 2002). The placenta and spleen cDNA inserts were unidirectionally cloned and sequenced from the 5′-end using the M13 reverse-48 primer (AGCGGATAACAATTTCACAC). Sequences were trimmed of vector, low-quality reads, selected for a minimum length of 200 bp and filtered for repeats using RepeatMasker (Smit and Green, 1999). Sequences were also filtered for contaminating sequences of viral, bacterial and mitochondrial origin. Clusters of ESTs were then created using CAP3 (Huang and Madan, 1999) default parameters, using 40 bp as the minimum size of the overlap between clones. After CAP3 assembly, all clusters and singlets containing sequences present on the 3800-gene array were removed from the clone set. New sequences were selected for the array using an approach that combined BLAST with evaluation of clone position in the transcript cluster. First, all sequences were analyzed by BLASTN against human UniGene (build 141) and checked for duplicates on the basis of human UniGene identification numbers of the best BLAST hits. Second, a representative clone was picked from each cluster with a UniGene identification number not represented on the 3800-gene array (clusters without UniGene hits were also used); clones with the longest and most high quality 3′ read available were selected. Finally, singletons with and without a human UniGene hit were added to the list of clones. A low level of redundancy was tolerated, particularly in cases when clones were identified with stronger sequence similarity scores to human UniGene clusters than the original clones used for the 3800 set. The total number of selected sequences for the microarray was 7872. Complete annotation of all the ESTs in the cattle placenta and spleen cDNA libraries used as a source to produce the microarray can be accessed at http://titan.biotec.uiuc.edu/cattle/cattle_project.htm.

Amplification of clone inserts, clean-up of PCR products and spotting of the microarray were performed as described previously (Band et al., 2002) with minor modifications. Amplification of inserts employed M13-FWD (GTTTTCCCAGTCACGACGTTG) and M13-REV (TGAGCGGATAACAATTTCACACAG) oligonucleotide primers (Hegde et al., 2000). After purification, PCR products were

redissolved in 3× SSC supplemented with 1.5 M betaine (Diehl et al., 2001). A row of control spots was placed in every grid of the array template. Positive controls include the endogenous housekeeping genes encoding beta actin (*ACTB*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and hypoxanthine phosphoribosyltransferase (*HPRT*). Exogenous spiking controls are the soybean genes chlorophyll ab binding protein (*CAB*), Rubisco small chain 1 (*RBS1*) and major latex protein (*MSG*). Negative controls are *Cot1* DNA, genomic DNA, spotting buffer, poly-A and H_2O. All PCR products were spotted in duplicate on the array. Robotic spotting of all cDNAs in duplicate was performed with a Cartesian Pixsys 5500 or GeneMachines OmniGrid 100 arrayer (Genomic Solutions, Ann Arbor, MI). Spot and printing quality were assessed on one slide by hybridizing a Cy3-labeled random nonamer (Operon, Alameda, Ca). The accuracy of the reracking, spotting and clone annotation was evaluated by resequencing the entire set of clones of the original 3800-gene microarray and sample sequencing eight clones per plate of the new clone set. Analysis of the sequence data revealed an error rate of 2% for the first set and 0% for the second set. Mislabeled clones were reannotated on the basis of the sequences obtained.

3. Functional annotation of microarray sequences

Sequence annotations have been updated several times since the initial round of clone selection for the micorarray. The following section describes the procedures used and a summary of the most recent annotation, conducted Spring, 2004. All 7872 sequences selected for the microarray were masked for repeats (RepeatMasker) and BLASTN similarity searches were conducted against human UniGene (build 166), mouse UniGene (build 135) and the human genome (build 34). Subsequently, the remaining sequences with an $E > e^{-5}$ were analyzed for similarity in predicted proteins by TBLASTX against the above-mentioned databases. In addition, BLASTN was performed against the bovine UniGene database (build 57) and the TIGR bovine database (build 9) in order to identify cattle-specific genes, EST clusters and annotations. For all searches, best hits were used to annotate the cattle sequences as putative orthologs. Previous comparative mapping studies have shown that such predictions are at least 95% accurate (Band et al., 2000). Perl scripts were used to annotate the cattle sequences with relevant information parsed from human UniGene and LocusLink (e.g., gene symbol, gene name, function, OMIM number, PubMed identification numbers) and obtain gene ontology (GO) annotations associated with human UniGene numbers (LocusLink, March 5, 2004). The GO flat files (March 5, 2004) were downloaded from http://www.geneontology.org/ (Ashburner et al., 2000) and used for GO annotation/classification of the sequences.

As described above, BLASTN of the 7872 sequences against human UniGene resulted in 6089 (77.4%) hits with $E < e^{-5}$. Subsequent BLASTN and TBLASTX analysis against mouse and human DNA sequence databases added 810 hits ($E < e^{-5}$). Thus, 6899 (87.4%) of cattle ESTs could be assigned a human or mouse UniGene identification number or human genome position, of which 5325 (77.2%) have unique UniGene identification numbers or human genome positions (Table 1). The remaining 973 sequences represent putatively novel genes of which a substantial number have an open reading frame greater than 100 amino acids (data not shown). Therefore, the 7872 cattle cDNA microarray may contain up to 6298 unique genes.

For the unique set of 5325 genes with UniGene identification numbers on the microarray, 3244 have GO annotation under the terms *molecular function* (1795), *biological process* (1809) and *cellular component* (1402) (Table 2). A significant number of genes

Table 1
Sequential BLAST search for annotation of the 7872-element cattle cDNA microarray

Database search	No. of hits	No. of unique hits
Human UniGene (BLASTN)	6089	4896
Mouse UniGene (BLASTN)	86	78
Human genome (BLASTN)	300	108
Human UniGene (TBLASTX)	326	212
Mouse UniGene (TBLASTX)	10	9
Human genome (TBLASTX)	88	22
Putative novel genes (no BLAST hit)	973	973
Total no. of genes (includes redundancy)	7872	6298

Table 2
Gene ontology (GO) annotations of the 3244 unique cattle genes on the 7872-element cDNA microarray

GO name[a]	GO term	No. of unique genes
Molecular function	0003674	1795
Antioxidant	0016209	2
Apoptosis regulator	0016329	16
Binding	0005488	841
Calcium ion binding	0005509	26
Carbohydrate binding	0030246	3
Drug binding	0008144	6
Glycosaminoglycan binding	0005539	6
Heavy metal binding	0005505	30
Lipid binding	0008289	18
Nucleic acid binding	0003676	421
DNA binding	0003677	266
Transcription factor	0003700	181
Nuclease	0004518	15
RNA binding	0003723	148
mRNA binding	0003729	31
Translation factor, nucleic acid binding	0008135	23
Nucleotide binding	0000166	131
Oxygen binding	0019825	5
Peptide binding	0042277	3
Protein binding	0005515	251
Calmodulin binding	0005516	8
Cytokine binding	0019955	8
Cytoskeletal protein binding	0008092	43
Transcription factor binding	0008134	95
Transcription cofactor	0003712	89
Receptor binding	0005102	59
Cell adhesion molecule	0005194	53
Calcium-dependent cell adhesion molecule	0008014	3
Cell adhesion receptor	0004895	16
Membrane-associated guanylate kinase	0004384	2
Chaperone	0003754	33
Co-chaperone	0003767	5
Heat shock protein	0003773	12
Defense/immunity protein	0003793	28
Antiviral response protein	0003800	10
Complement activity	0003811	8
Enzyme	0003824	660
Helicase	0004386	25
Histone deacetylase	0004407	4
Hydrolase	0016787	284
Isomerase	0016853	21
Kinase	0016301	120
Ligase	0016874	34
Lyase	0016829	17
Oxidoreductase	0016491	98
Phosphatase	0016302	46
Small protein conjugating enzyme	0008639	23
Transferase	0016740	196
Enzyme regulator	0030234	86
Enzyme activator	0008047	31
Enzyme inhibitor	0004857	51
Kinase regulator	0019207	12
Phosphatase regulator	0019208	8
Motor	0003774	14
Signal transducer	0004871	281
Receptor	0004872	139
Receptor binding	0005102	59
Receptor signaling protein	0005057	65
Receptor-associated protein	0016962	5
Structural molecule	0005198	106
Structural constituent of cytoskeleton	0005200	24
Structural constituent of muscle	0008307	9
Structural constituent of ribosome	0003735	47
Transcription regulator	0030528	232
Transcription cofactor	0003712	89
Transcription co-activator	0003713	48
Transcription co-repressor	0003714	38
Transcription factor	0003700	181
RNA polymerase II transcription factor	0003702	71
Transcription elongation factor	0003711	4
Translation regulator	0045182	23
Translation factor, nucleic acid binding	0008135	23
Transporter	0005215	165
Amino acid transporter	0015171	5
Auxiliary transport protein	0015457	4
Carrier	0005386	54
Channel/pore class transporter	0015267	18
Electron transporter	0005489	45
Intracellular transporter	0005478	6
Ion transporter	0015075	48
Lipid transporter	0005319	7
Protein transporter	0008565	4
Biological process	0008150	1809
Behavior	0007610	6
Cell communication	0007154	593
Cell adhesion	0007155	64
Cell–cell signaling	0007267	72
Response to external stimulus	0009605	233
Signal transduction	0007165	385
Cell growth and/or maintenance	0008151	1387
Cell cycle	0007049	156

Table 2 (*Continued*)

GO name[a]	GO term	No. of unique genes
DNA replication and chromosome cycle	0000067	30
M phase	0000279	31
Mitotic cell cycle	0000278	66
Regulation of cell cycle	0000074	104
Cell motility	0006928	89
Cell organization and biogenesis	0016043	80
Cell proliferation	0008283	140
Cellular morphogenesis	0000902	22
Colony morphology	0007149	13
Regulation of cell shape	0008360	8
Homeostasis	0019725	14
Membrane fusion	0006944	9
Metabolism	0008152	984
Alcohol metabolism	0006066	29
Amine metabolism	0009308	30
Amino acid and derivative metabolism	0006519	26
Biosynthesis	0009058	144
Carbohydrate metabolism	0005975	57
Catabolism	0009056	110
Coenzymes and prosthetic group metabolism	0006731	7
Electron transport	0006118	18
Energy pathways	0006091	60
Lipid metabolism	0006629	88
Nitrogen metabolism	0006807	11
Nucleobase, nucleoside, nucleotide and nucleic acid metabolism	0006139	393
Organic acid metabolism	0006082	54
Oxygen and reactive oxygen species metabolism	0006800	16
Phosphorus metabolism	0006793	102
Protein metabolism	0019538	403
Sulfur metabolism	0006790	5
Response to stress	0006950	156
Response to DNA damage	0006974	14
Response to oxidative stress	0006979	14
Response to pest/pathogen/ parasite	0009613	104
Response to wounding	0009611	65
Transport	0006810	193
Cytoplasmic transport	0016482	28
Hydrogen transport	0006818	9
Ion transport	0006811	32
Lipid transport	0006869	4
Protein transport	0015031	88
Vesicle mediated transport	0016192	86
Death	0016265	82
Cell death	0008219	82
Programmed cell death	0012501	80
Apoptosis	0006915	79
Induction of programmed cell death	0012502	31
Development	0007275	219
Genetic transfer	0009292	4
Morphogenesis	0009653	148
Cellular morphogenesis	0000902	22
Organogenesis	0009887	128
Post-embryonic morphogenesis	0009886	3
Pattern specification	0007389	3
Post-embryonic development	0009791	3
Regulation of gene expression, epigenetic	0040029	4
Reproduction	0000003	18
Actin cytoskeleton reorganization	0007012	13
Embryogenesis and morphogenesis	0007345	34
Histogenesis and organogenesis	0007397	22
Oncogenesis	0007048	117
Small molecule transport	0006832	43
Viral replication	0008166	10
Virulence	0009406	6
Physiological processes	0007582	78
Circulation	0008015	21
Digestion	0007586	5
Excretion	0007588	6
Hemostasis	0007599	12
Nutritional response pathway	0007584	7
Pathogenesis	0009405	21
Pregnancy	0007565	5
Cellular component	0005575	1402
Cell	0005623	1293
Cell fraction	0000267	190
Insoluble fraction	0005626	3
Membrane fraction	0005624	136
Soluble fraction	0005625	57
Intracellular	0005622	1193
Cell cortex	0005938	4
Chromosome	0005694	20
Cytoplasm	0005737	653
Nucleus	0005634	364
Plasma membrane	0005886	275
Proton-transporting ATP synthase complex	0045259	4
Respiratory chain complex I	0045271	19
Ribonucleoprotein complex	0030529	72
Ubiquitin ligase complex	0000151	5
Membrane	0016020	434
Endomembrane system	0012505	42
Inner membrane	0019866	39
Integral to membrane	0016021	273

Table 2 (*Continued*)

GO name[a]	GO term	No. of unique genes
Mitochondrial membrane	0005740	47
Plasma membrane	0005886	275
Extracellular	0005576	116
Extracellular matrix	0005578	40
Basement membrane	0005604	13
Collagen	0005581	9
Extracellular space	0005615	63

[a] Indentation denotes a subcategory of the hierarchical term.

encoding transcription factors (GO:0003700), signal transducers (GO:0004871), cell cycle regulators (GO:0000074) and programmed cell death (GO:0012501) are represented on the microarray (Table 2). The exhaustive annotation of the 7872 cDNA array that was produced using multiple-species databases and TBLASTX for identification of divergent orthologs provides a fundamental tool for understanding the large experimental datasets collected in our studies. Annotation of all sequences on the array, including sequence accession number, gene symbol, *E*-value of best-hit and GO terms can be found on at http://cagst.animal.uiuc.edu/microarray/.

4. Experiments being conducted using the 7872 cDNA microarray

The microarray described in this report has been used in several large-scale research programs during the past 2 years. Collaborative projects were chosen to address important problems in animal breeding and genetics, reproduction, development, nutrition and disease resistance. Major components of these studies are now completed or are nearing completion. A brief summary of each of these projects is given below.

4.1. Cloning efficiency and developmental genomics

Approximately 40% of all pregnancies derived from artificial insemination (AI) or in vitro fertilization (IVF), and greater than 90% of pregnancies derived from nuclear transfer (NT) cloning, terminate prematurely. Of these losses, 80% are attributable to a dysfunctional placenta (Heyman et al., 2002). Critical time periods have been identified in which embryonic or fetal survival rate decline (Heyman et al., 2002). These windows of development, especially around days 7, 18–21, 35 and 50–70 of gestation, represent ideal time points for functional genomic analysis.

A multi-institution, multi-national collaboration has been established to study the problem of cloning efficiency and the phenotypic abnormalities associated with the cloning process, such as large offspring syndrome (Young et al., 1998; Chavatte-Palmer et al., 2002). The collaboration involves the University of Illinois, Institut National de la Recherche Agronimique (INRA), France (Jean Paul Renard, Isabelle Hue and co-workers), The University of Connecticut at Storrs (Jerry Yang, Cindy Tian and co-workers) and USDA Beltsville Agricultural Research Center (Kurt Zuelke), with funding provided by USDA-Agricultural Research Service. To date, we have studied the pre-implantation period at day 7 and the post-implantation period at term. Individual 7-day embryos derived from AI, IVF and NT were profiled using the 7872 element cDNA array. To perform expression profiling on the embryos, a linear mRNA amplification protocol (Baugh et al., 2001) was used to generate ample material for analysis. We also compared the expression patterns obtained with embryos to the original donor cell line. Our goal is to use gene expression patterns to (i) assess the extent of nuclear reprogramming of donor cell lines, (ii) to analyze and interpret effects of cloning and in vitro culture on gene expression in the embryo and (iii) to correlate these effects with embryo and fetus survival.

In addition to the analysis of 7-day embryos, gene expression profiling of trophoblast, placentomes and fetal organs will be performed at different times of development. Placentomes collected from AI, IVF and NT pregnancies after caesarean section were profiled for gene expression using the 7872-element cDNA array. Furthermore, phenotypes of all calves were recorded in order to identify specific gene expression profiles associated with phenotypic abnormalities (including large offspring syndrome). The main questions to be answered in the analysis of placentomes during different stages of pregnancy are whether specific genes and pathways are altered by NT, IVF or embryo culture conditions, and if these changes are ‘‘stable’’ in embryos and fetuses that survive the IVF and cloning processes. Our analysis to date has demonstrated clear differences in gene

expression among the groups. These results may have great significance in the debate over the ''normalcy'' of calves derived from the NT process.

4.2. Analysis of tax-regulated cellular pathways involved in BLV infection

Bovine leukemia virus (BLV) is an oncogenic B-lymphotropic retrovirus and the etiologic agent of enzootic bovine leukosis (Sagata et al., 1985; Kettmann et al., 1994). Even though the process of leukemogenesis caused by retroviruses is still unclear, the BLV-*tax* gene has been demonstrated to play a key role in the pathogenesis of diseases caused by the primate T-cell lymphotropic viruses (PTLV-1, -2 and -3) and BLV (Smith and Greene, 1991; Philpott and Buehring, 1999). Moreover, Tax protein appears essential for human T-cell lymphotrophic virus type 1 (HTLV-1)-associated immortalization and transformation of T-cells (Tanaka et al., 1990). Tax is required for viral replication in vivo and activates transcription of not only viral genes, but also several cellular genes, including those for various cytokines thought to affect viral spread and disease progression (Amills et al., 2002; Twizere et al., 2003). The main objective of our studies is to understand the genomic effects of BLV-Tax expression in host B cells and to relate those changes to the development of preneoplastic and neoplastic disease. To approach this problem, we have conducted transcript profiling of BL3°, a transformed but uninfected bovine B lymphoblastoid cell line and BL3*, a BLV-infected derivative of BL3° (Romano et al., 1989). In a preliminary study, approximately 270 genes were shown to be differentially expressed in BL3* in comparison to its uninfected parental cell line BL3° (Everts et al., 2002), many of which were also found to be differentially expressed in HTLV-1-infected cell lines (Pise-Masison et al., 2002). We are now using RNA interference technology (Elbashir et al., 2001) to knock down the expression of the BLV-*tax* gene in BL3* cell line in order to dissect *tax*-associated effects from possible effects associated with derivation of this cell line. The effects of the *tax* mRNA knock-down on host gene expression will be evaluated by microarray analysis of RNA obtained from BL3* before and after introduction of the *tax*-specific siRNA. We anticipate that results of our experiments will provide further understanding of the importance of BLV-*tax* in B-cell transformation and activation of the host immune response (Nomura et al., 2004).

4.3. Nutritional genomics

Nutritional or management limitations during the dry period of lactation may impede the ability of the dairy cow to reach maximal milk production (Drackley, 1999). The periparturient, or *transition period* (3 weeks before to 3 weeks after parturition), may be the most critical phase of the lactation cycle (Drackley, 1999). We have begun a collaborative effort with James Drackley (University of Illinois) funded by the USDA-National Research Initiative to unravel patterns of gene expression in liver of dairy cows in response to physiological state, plane of nutrition and metabolic disorders such as ketosis. A longitudinal assessment of hepatic gene expression in cows with adequate, excess or restricted nutrition prepartum has been obtained in liver biopsied at −65, −30, −14, +1, +14, +28 and +49 days relative to parturition (Loor et al., 2004d, 2004e). Another experiment (Loor et al., 2004a) examined hepatic gene expression in cows subjected to an acute feed restriction (50% of ad libitum intake at mid-lactation) in order to determine differential effects of inadequate nutrition on mRNA abundance without the potential confounding of the hormonal environment characteristic of the peripartum period. A pilot study of gene expression in mammary tissue collected simultaneously with mammary and liver biopsies (Loor et al., 2004b) during the periparturient period (−14, +1 and +14 days) has also been conducted. Expression profiles in liver from cows induced to develop primary ketosis early postpartum at +10 to +14 days after parturition (Loor et al., 2004c), revealed a unique set of genes that will be used to delineate previously unknown hepatic adaptations due to ketosis.

4.4. Transcription profiling of genetic merit

During the past half century, animal breeding has been enormously effective at increasing the productivity and profitability of the livestock and poultry

industries. Traditional animal breeding uses phenotypic records of an individual and its relatives to predict the genetic merit of that individual and its offspring. Genetic markers have been introduced to control for deleterious mutations as well as to detect those genes affecting quantitative traits (quantitative trait loci or QTL) and for ''marker-assisted-selection'' of genetically elite animals. However, whole-genome scanning techniques are expensive and the resolution of QTL maps is still generally quite low, making implementation of marker-assisted breeding schemes difficult for many economically important traits. The low resolution of QTL mapping and the uncertainty of the comparative positional candidate approach suggest a need for new approaches to identify the actual genes controlling complex traits.

We have recently completed a study in collaboration with Sandra Rodriguez-Zas (University of Illinois) that tested the hypothesis that selection for increased milk production influences systemic gene expression profiles. Peripheral blood leukocytes and liver tissue biopsied from heifers with extreme values of predicted transmitting ability (PTA) for milk yield were profiled for gene expression and analyzed for differences between the high and low PTA groups. Expression levels of ~50 genes in peripheral blood leukocytes were differentially expressed and a smaller subset of 14 genes accurately predicted animals in the high or low groups. Several of the differentially expressed genes fall into regions, where QTL have been mapped. Similarly, recent results with rats showed that QTL for hypertension can be identified by gene expression profiling (Aitman et al., 1999). These exciting results indicate that gene expression profiling of peripheral blood leukocytes of young heifers can be used to predict their milk production levels as adults. Furthermore, many of the identified genes are candidates for improving milk production and can be considered as logical drug targets, as pharmaceuticals or for transgenic manipulation. If the method proves to be predictive of PTA for production traits of breeding bulls, it will have the potential to radically transform dairy and beef breeding strategies. We have termed the strategy of using gene expression profiles for genetic improvement ''phenomic selection'' and plan follow-up validation studies on animals of different sex and age in different herds.

4.5. Gene expression profiling of normal tissues

Gene expression patterns in healthy and diseased tissues have proven to be a valuable tool for disease diagnosis, elucidating molecular mechanisms of disease pathogenesis and for gene annotation. One of our early efforts with the 7000-series microarray was to profile gene expression patterns in 17 different tissues collected by vivisection from a 1-week-old male Jersey calf (Band et al., 2003). Cluster analysis revealed a large number of tissue-specific signatures consisting of hundreds of genes. In addition, tissues with related functions, such as those associated with immune responses, the digestive system and the central nervous system, showed distinctive gene expression patterns. Expression levels of previously unannotated and ''novel'' transcripts (Lewin et al., 2004) were found to group in tissue-specific clusters thus demonstrating the utility of microarray technology for cattle-specific functional annotation of gene expression. The gene expression patterns of normal tissues should serve as a useful reference for disease diagnosis, prognosis, pathogenesis and therapy.

4.6. Comparative genomic analysis of the bovine host response to intracellular zoonotic pathogens

Relatively little is known about the comparative pathogenesis of many of the common zoonotic diseases caused by intracellular pathogens. The U.S. Department of Homeland Security has recently awarded a major grant to a consortium of public universities headed by Texas A&M University (Neville P. Clarke, PI) to study essential pathways of host responses to several intracellular zoonotic pathogens. In collaboration with Garry Adams at Texas A&M, transcription profiling will be performed on tissues collected from cattle at different time points after exposure to *Salmonella enterica* Typhimurium, *Brucella abortus*, *Cryptosporidium parvum* and *Mycobacterium avium* subspecies *paratuberculosi*s. The objective is to identify specific genes in common pathway(s) essential for these intracellular pathogens to colonize and exploit the host resulting in morbidity and mortality. Identification of the critical ''disease dependent'' host gene(s) will be the basis for generating targeted genetic disease resistance through

marker-assisted selection, cloning and engineering knock-out or knock-in transgenic cattle.

5. Future improvements and innovations

The rapid advancements in cattle EST and genome sequencing, in particular the doubling in GenBank of EST entries during the past year (432,774 on May 19, 2004), provide the raw material for dramatic improvements in the current technology for cattle functional genomics. Furthermore, the forthcoming draft cattle genome sequence will add an invaluable resource for development of technologies for whole genome transcript profiling. With the low cost of long (70-mer) oligo synthesis and the overall cost advantages of spotted oligos over spotted cDNA arrays, there is increasing interest in deploying the spotted oligo platform for cattle functional genomics. We have performed a comparison study using ~200 long oligos and cDNAs and found the correlation to be ~0.7 for the same RNA (BL3* cell line; unpublished data). This is in general agreement with similar studies in other species (Kuo et al., 2002; Carter et al., 2003; Wang et al., 2003). While the nature of the discrepancies between oligos and cDNAs remains unclear for all cases, alternative splicing, paralogs and design errors probably account for most of the differences. To be certain that the discrepancies are not platform-dependent, more sensitive methods, such as quantitative RT-PCR, can be deployed.

The reasonably good correlation between results obtained with cDNA and 70-mer oligos convinced us to switch to the oligo platform; we are currently producing a 13,000-gene 70-mer oligo array for our future work. The additional 6000 genes we have added to the oligo array are contained within a new collection of cDNA clones end-sequenced from cDNA libraries created at the University of Illinois W. M. Keck Center for Comparative and Functional Genomics. These new ESTs (~20,000) were derived from cDNA libraries created primarily from bovine fetal tissues (in collaboration with Jerry Yang and Cindy Tian, University of Connecticut at Storrs; Jean Paul Renard and Isabelle Hue, INRA, France) and all have been deposited in GenBank. Having the cDNA clones for all 13,000 unique genes in our possession may be important for confirmatory and follow-up functional studies. We plan to expand the oligo design in about 1 year to accommodate the cattle genome sequence and the large number of additional EST clusters that can be annotated to it (e.g., there are currently >35,000 TIGR clusters of cattle ESTs). At that time, we may consider changing from the spotted oligo platform to other platforms that allow higher density spotting. We are also aware that Affymetrix as well as other groups around the world plan to produce microarrays for bovine gene expression studies. The future will likely belong to the efficient, as cost *will* matter in the livestock functional genomics arena. Furthermore, any efforts to create microarrays for broad usage must also include a strong bioinformatics component that will take into account comparative data as well as the anticipated fast-growing database(s) containing cattle-specific annotation. The relatively large number of lineage associated expansions of gene families, such as the *PAG*s (Xie et al., 1997) and *MHCLA/ULBPs* (Larson et al., 2003), and accounting for splice variants and non-coding RNAs, will be particularly important as the questions being asked with the technology become more refined. Profiling of the entire cattle transcriptome will be possible in the near future, but fully understanding the results of even the simplest studies will take more sophisticated data mining technologies than currently available.

Acknowledgments

The creation of cattle microarrays was made possible in part by grants from the USDA-National Research Initiative (AG99-35205-8534) and the USDA-Agricultural Research Service (AG 58-1265-2-020). Funding sources for individual projects were cited in the text.

References

Aitman, T.J., Glazier, A.M., Wallace, C.A., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., al Majali, K.M., Trembling, P.M., Mann, C.J., Shoulders, C.C., Graf, D., St Lezin, E., Kurtz, T.W., Kren, V., Pravenec, M., Ibrahimi, A., Abumrad, N.A., Stanton, L.W., Scott, J., 1999. Identification of *Cd36* (*Fat*) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. Nat. Genet. 21, 76–83.

Amills, M., Ramiya, V., Norimine, J., Olmstead, C.A., Lewin, H.A., 2002. Reduced IL-2 and IL-4 mRNA expression in $CD4^+$ T cells

from bovine leukemia virus-infected cows with persistent lymphocytosis. Virology 304, 1–9.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortiums. Nat. Genet. 25, 25–29.

Band, M.R., Everts, R.E., Liu, Z.L., Morin, D.E., Peled, J.U., Rodriguez-Zas, S., Lewin, H.A., 2003. Gene expression profiling of 17 cattle tissues reveals unique patterns related to tissue function. In Final Abstracts Guide, Plant and Animal Genome X, San Diego, USA, p 260.

Band, M.R., Larson, J.H., Rebeiz, M., Green, C.A., Heyen, D.W., Donovan, J., Windish, R., Stiening, C.M., Mahyuddin, P., Womack, J.E., Lewin, H.A., 2000. An ordered comparative map of the cattle and human genomes. Genome Res. 10, 1359–1368.

Band, M.R., Olmstead, C., Everts, R.E., Liu, Z.L., Lewin, H.A., 2002. A 3800 gene microarray for cattle functional genomics: comparison of gene expression in spleen, placenta and brain. Anim. Biotechnol. 13, 163–172.

Baugh, L.R., Hill, A.A., Brown, E.L., Hunter, C.P., 2001. Quantitative analysis of mRNA amplification by in vitro transcription. Nucleic Acids Res. 29, E29.

Burton, J.L., Madsen, S.A., Yao, J., Sipkovsky, S.S., Coussens, P.M., 2001. An immunogenomics approach to understanding periparturient immunosuppression and mastitis susceptibility in dairy cows. Acta Vet. Scand. 42, 407–424.

Carter, M.G., Hamatani, T., Sharov, A.A., Carmack, C.E., Qian, Y., Aiba, K., Ko, N.T., Dudekula, D.B., Brzoska, P.M., Hwang, S.S., Ko, M.S., 2003. In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. Genome Res. 13, 1011–1021.

Chavatte-Palmer, P., Heyman, Y., Richard, C., Monget, P., Lebourhis, D., Kann, G., Chilliard, Y., Vignon, X., Renard, J.P., 2002. Clinical, hormonal, and hematologic characteristics of bovine calves derived from nuclei from somatic cells. Biol. Reprod. 66, 1596–1603.

Coussens, P.M., Colvin, C.J., Wiersma, K., Abouzied, A., Sipkovsky, S., 2002. Gene expression profiling of peripheral blood mononuclear cells from cattle infected with *Mycobacterium paratuberculosis*. Infect. Immun. 70, 5494–5502.

Diehl, F., Grahlmann, S., Beier, M., Hoheisel, J.D., 2001. Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. Nucleic Acids Res. 29, E38.

Drackley, J.K., 1999. ADSA Foundation Scholar Award. Biology of dairy cows during the transition period: the final frontier? J. Dairy Sci. 82, 2259–2273.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T., 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411, 494–498.

Everts, R.E., Band, M.R., Liu, Z.L., Peled, J.U., Kumar, C.G., Bari, A., Liu, L., Lewin, H.A., 2002. Transcript profiling of bovine leukemia virus (BLV)-infected and -uninfected cell lines using a cDNA microarray containing 7653 cattle genes. In Final Abstracts Guide, International Society of Animal Genetics, Göttingen, Germany, p 55.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S., 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531–537.

Hamatani, T., Carter, M.G., Sharov, A.A., Ko, M.S., 2004. Dynamics of global gene expression changes during mouse preimplantation development. Dev. Cell 6, 117–131.

Hansson, A., Hance, N., Dufour, E., Rantanen, A., Hultenby, K., Clayton, D.A., Wibom, R., Larsson, N.G., 2004. A switch in metabolism precedes increased mitochondrial biogenesis in respiratory chain-deficient mouse hearts. Proc. Natl. Acad. Sci. U.S.A. 101, 3136–3141.

Hashizume, K., Ishiwata, H., Kizaki, K., Yamada, O., Takahashi, T., Imai, K., Patel, O.V., Akagi, S., Shimizu, M., Takahashi, S., Katsuma, S., Shiojima, S., Hirasawa, A., Tsujimoto, G., Todoroki, J., Izaike, Y., 2002. Implantation and placental development in somatic cell clone recipient cows. Cloning Stem Cells 4, 197–209.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N., Quackenbush, J., 2000. A concise guide to cDNA microarray analysis. Biotechniques 29, 548–560.

Herath, C.B., Shiojima, S., Ishiwata, H., Katsuma, S., Kadowaki, T., Ushizawa, K., Imai, K., Takahashi, T., Hirasawa, A., Tsujimoto, G., Hashizume, K., 2004. Pregnancy-associated changes in genome-wide gene expression profiles in the liver of cow throughout pregnancy. Biochem. Biophys. Res. Commun. 313, 666–680.

Heyman, Y., Chavatte-Palmer, P., Lebourhis, D., Camous, S., Vignon, X., Renard, J.P., 2002. Frequency and occurrence of late-gestation losses from cattle cloned embryos. Biol. Reprod. 66, 6–13.

Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.

Ishiwata, H., Katsuma, S., Kizaki, K., Patel, O.V., Nakano, H., Takahashi, T., Imai, K., Hirasawa, A., Shiojima, S., Ikawa, H., Suzuki, Y., Tsujimoto, G., Izaike, Y., Todoroki, J., Hashizume, K., 2003. Characterization of gene expression profiles in early bovine pregnancy using a custom cDNA microarray. Mol. Reprod. Dev. 65, 9–18.

Kettmann, R.A., Burny, A., Callebaut, I., Droogmans, L., Mammerickx, M., Willems, L., Portetelle, D., 1994. Bovine leukemia virus. In: Levy, J.A. (Ed.), The Retroviridae. Plenum Press, New York, pp. 39–81.

Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L., Kohane, I.S., 2002. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 18, 405–412.

Larson, J.H., Rebeiz, M.J., Stiening, C.M., Windish, R.L., Beever, J.E., Lewin, H.A., 2003. MHC class I-like genes in cattle, *MHCLA*, with similarity to genes encoding NK cell stimulatory ligands. Immunogenetics 55, 16–22.

Lewin, H.A., Larson, J.H., Kumar, C.G., 2004. Comparative mammalian genomics and adaptive evolution: divergent

homologs and novel genes in the cattle genome. In: Wasser, S.P. (Ed.), Evolutionary Theory and Processes: Modern Horizons. Kluwer Academic Publishers, pp. 139–152.

Loor, J.J., Carlson, D.B., Everts, R.E., Rodriguez-Zas, S., Lewin, H.A., Drackley, J.K., 2004a. Gene expression profiles in liver of dairy cows in response to feed restriction using a bovine cDNA microarray. J. Dairy Sci. 87, T136.

Loor, J.J., Dann, H.M., Everts, R.E., Rodriguez-Zas, S., Lewin, H.A., Drackley, J.K., 2004b. Mammary and hepatic gene expression analysis in peripartal dairy cows using a bovine cDNA microarray. J. Dairy Sci. 87, T134.

Loor, J.J., Dann, H.M., Morin, D.E., Everts, R.E., Rodriguez-Zas, S., Lewin, H.A., Drackley, J.K., 2004c. Early postpartum ketosis in dairy cows and hepatic gene expression profiles using a bovine cDNA microarray. In: Proceedings of the 12th International Conference on Production Diseases in Farm Animals. p. 77.

Loor, J.J., Drackley, J.K., Dann, H.M., Everts, R.E., Rodriguez-Zas, S., Lewin, H.A., 2004d. Gene expression patterns in liver of dairy cows from dry-off through early lactation using a bovine cDNA microarray. FASEB J. 18, 4847.

Loor, J.J., Janovick, N.A., Dann, H.M., Everts, R.E., Rodriguez-Zas, S., Lewin, H.A., Drackley, J.K., 2004e. Microarray analysis of hepatic gene expression from dry-off through early lactation in dairy cows fed at two intakes during the dry period. J. Dairy Sci. 87, T133.

Lotinun, S., Sibonga, J.D., Turner, R.T., 2003. Triazolopyrimidine (trapidil), a platelet-derived growth factor antagonist, inhibits parathyroid bone disease in an animal model for chronic hyperparathyroidism. Endocrinology 144, 2000–2007.

Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett Jr., D.E., Hartwell, L.H., Brown, P.O., Friend, S.H., 1998. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat. Med. 4, 1293–1301.

Nomura, M., Ohashi, T., Nishikawa, K., Nishitsuji, H., Kurihara, K., Hasegawa, A., Furuta, R.A., Fujisawa, J., Tanaka, Y., Hanabuchi, S., Harashima, N., Masuda, T., Kannagi, M., 2004. Repression of tax expression is associated both with resistance of human T-cell leukemia virus type 1-infected T-cells to killing by tax-specific cytotoxic T lymphocytes and with impaired tumorigenicity in a rat model. J. Virol. 78, 3827–3836.

Philpott, S.M., Buehring, G.C., 1999. Defective DNA repair in cells with human T-cell leukemia/bovine leukemia viruses: role of tax gene. J. Natl. Cancer Inst. 91, 933–942.

Pise-Masison, C.A., Radonovich, M., Mahieux, R., Chatterjee, P., Whiteford, C., Duvall, J., Guillerm, C., Gessain, A., Brady, J.N., 2002. Transcription profile of cells infected with human T-cell leukemia virus type I compared with activated lymphocytes. Cancer Res. 62, 3562–3571.

Reverter, A., Byrne, K.A., Brucet, H.L., Wang, Y.H., Dalrymple, B.P., Lehnert, S.A., 2003. A mixture model-based cluster analysis of DNA microarray gene expression data on Brahman and Brahman composite steers fed high-, medium-, and low-quality diets. J. Anim. Sci. 81, 1900–1910.

Romano, M.J., Stewart, J.A., Lewin, H.A., 1989. Phenotypic characterization of bovine lymphoblastoid cell lines. Vet. Immunol. Immunopathol. 23, 293–307.

Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y., Ikawa, Y., 1985. Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. Proc. Natl. Acad. Sci. U.S.A. 82, 677–681.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.

Smit, A.F.A, Green, P. Repeatmasker at http://ftp.genome.washington.edu/. 1999.

Smith, M.R., Greene, W.C., 1991. Molecular biology of the type I human T-cell leukemia virus (HTLV-I) and adult T-cell leukemia. J. Clin. Invest. 87, 761–766.

Suchyta, S.P., Sipkovsky, S., Halgren, R.G., Kruska, R., Elftman, M., Weber-Nielsen, M., Vandehaar, M.J., Xiao, L., Tempelman, R.J., Coussens, P.M., 2003a. Bovine mammary gene expression profiling using a cDNA microarray enhanced for mammary-specific transcripts. Physiol. Genomics 16, 8–18.

Suchyta, S.P., Sipkovsky, S., Kruska, R., Jeffers, A., McNulty, A., Coussens, M.J., Tempelman, R.J., Halgren, R.G., Saama, P.M., Bauman, D.E., Boisclair, Y.R., Burton, J.L., Collier, R.J., DePeters, E.J., Ferris, T.A., Lucy, M.C., McGuire, M.A., Medrano, J.F., Overton, T.R., Smith, T.P., Smith, G.W., Sonstegard, T.S., Spain, J.N., Spiers, D.E., Yao, J., Coussens, P.M., 2003b. Development and testing of a high-density cDNA microarray resource for cattle. Physiol. Genomics 15, 158–164.

Takahashi, M., Rhodes, D.R., Furge, K.A., Kanayama, H., Kagawa, S., Haab, B.B., Teh, B.T., 2001. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. Proc. Natl. Acad. Sci. U.S.A. 98, 9754–9759.

Tanaka, A., Takahashi, C., Yamaoka, S., Nosaka, T., Maki, M., Hatanaka, M., 1990. Oncogenic transformation by the *tax* gene of human T-cell leukemia virus type I in vitro. Proc. Natl. Acad. Sci. U.S.A. 87, 1071–1075.

Twizere, J.C., Kruys, V., Lefèbvre, L., Vanderplasschen, A., Collete, D., Debacq, C., Lai, W.S., Jauniaux, J.C., Bernstein, L.R., Semmes, O.J., Burny, A., Blackshear, P.J., Kettmann, R., Willems, L., 2003. Interaction of retroviral tax oncoproteins with tristetraprolin and regulation of tumor necrosis factor-α expression. J. Natl. Cancer Inst. 95, 1846–1859.

Wang, H.Y., Malek, R.L., Kwitek, A.E., Greene, A.S., Luu, T.V., Behbahani, B., Frank, B., Quackenbush, J., Lee, N.H., 2003. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. Genome Biol. 4, R5.

Whitfield, C.W., Cziko, A.M., Robinson, G.E., 2003. Gene expression profiles in the brain predict behavior in individual honey bees. Science 302, 296–299.

Xie, S., Green, J., Bixby, J.B., Szafranska, B., DeMartini, J.C., Hecht, S., Roberts, R.M., 1997. The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase subfamily consisting of many trophoblast-expressed genes. Proc. Natl. Acad. Sci. U.S.A. 94, 12809–12816.

Yao, J., Burton, J.L., Saama, P., Sipkovsky, S., Coussens, P.M., 2001. Generation of EST and cDNA microarray resources for the study of bovine immunobiology. Acta Vet. Scand. 42, 391–405.

Young, L.E., Sinclair, K.D., Wilmut, I., 1998. Large offspring syndrome in cattle and sheep. Rev. Reprod. 3, 155–163.